Supporting information for :

Insights into Structure and Redox Potential of Lignin Peroxidase from QM/MM Calculations

*Ludovic Castro, L. Ellis Crawford, Archford Mutengwa, Jan P. Götze and Michael Bühl**

University of St Andrews, School of Chemistry, North Haugh, St Andrews, Fife, KY16 9ST, Scotland, United Kingdom

Computational Details

The setup of the system

The initial geometries of Cpd I and Cpd II of LiP were constructed from the experimental crystal structure of the major LiP isozyme from *Phanerocheate chrysosporium* (PDB file 1LGA, resolution of 2.03 Å).[1] The refined model consists of two LiP molecules in the asymmetric unit, but only the first biological assembly was used because the interaction between both monomers is weak. The oxygen of water molecule 471 interacting with iron was replaced by an oxo group in order to model Cpd I and Cpd II. The distance between the oxygen and the iron atoms has been adjusted in order to correspond to a double Fe=O bond. After adding H atoms and neutralizing the system through protonation of suitable titratable residues, a hydration sphere of 40 Å around the central Fe was added. The complete system consists of 25228 atoms, including 6715 water molecules. Afterwards, molecular dynamics simulation of 1 ns (T = 300 K) was carried out on the entire enzyme, during which the coordinates of the heme unit and the coordinating His-176 were kept fixed (because of the lack of suitable force-field parameters for the heme part). In order to observe the impact of different enzyme conformations on the redox potential calculations, 10 different snapshots (after every 100 ps between 100 - 1000 ps of equilibration) were chosen as starting points for the QM/MM optimizations.

The glucosamine unit N-linked to Asn-257 was replaced by a hydrogen atom because its parameters are not as well defined as protein residues in the CHARMM force field. Its

removal should not have an important impact on the calculated properties since it is a neutral species located on the surface of the enzyme (located at 24 Å from Fe). The protonation states of the titratable residues were determined using PROPKA[2] in combination with visual inspection. In order to favor the formation of hydrogen bonds with their local environment, the histidines, His-30, His-47, His-82, His-149, His-176, His-310 and His-341 were protonated at the δ position, while His-39, His-101 and His-239 were protonated at the ε position. One of the propionate groups of the heme is protonated at O2A in order to form a hydrogen bond with the aspartate Asp-183. Arginines (Arg) and lysines (Lys) were used as positively charged. Aspartates, Asp-11, Asp-19, Asp-22, Asp-23, Asp-88, Asp-152, Asp-185, Asp-263, Asp-278, Asp-287 and Asp-312 were protonated, as well as glutamates, Glu-57, Glu-89, Glu-163, Glu-166, Glu-168, Glu-226, Glu-232, Glu-314 and Glu-319, in order to neutralize the global charge of the enzyme.[3] The missing hydrogen atoms were added *via* the HBUILD[4] module and optimized by the CHARMM force field as implemented in the CHARMM program[5] using the adopted basis Newton-Raphson method[6]. Then, a thick water sphere of 40 Å of radius was constructed around the enzyme using the standard tools in CHARMM. The following procedure was carried out to equilibrate the solvation sphere: 1) Addition of the water sphere and deletion of every water molecule located at less than 2.8 Å of another atom. 2) MM optimization for 200 steps. 3) Heating to 300 K. 4) Equilibration for 2 ps. The entire procedure was repeated until less than 100 water molecules were effectively added during step 1.

Calcium binding sites.

The distal calcium interacts with the carbonyl oxygen atoms of Asp-48 and Gly-66, the side-chain oxygen atom of Ser-70, one side chain carboxyl oxygen atom of Asp-48 and Asp-68, and two water molecules. Both side chain carboxyl oxygen atoms of Glu-78 form hydrogen bonds with one of these water molecules and with the OH group of Ser-70. Two more important H-bonds exist between the carbonyl oxygen of Glu-78 and the NH_2 group of Asn-84, and between the side chain oxygen of Asn-84 and the distal histidine His-47. Finally, Asp-48 immediately follows His-47. This succession of interactions shows that the distal calcium maintains the global integrity of the distal part of the enzyme. The chain of interactions creates a ring which can be written as Ca:Ser-70:Glu-78:Asn-84:His-47:Asp-48:Ca. The proximal calcium interacts with the carbonyl oxygen atoms of Ile-199, Ser-177 and Thr-196, the side chain oxygen atoms of Ser-177 and Thr-196, one side chain carboxyl oxygen atom of Asp-201, and both side chain carboxyl oxygen atoms of Asp-194. This cation is the most

tightly bound since water doesn't serve as a ligand, contrary to the distal calcium. Ser-177 follows immediately His-176, the proximal histidine in direct interaction with the iron of the active site. This also shows the importance of the proximal calcium for the integrity of the proximal part of the enzyme.

The QM/MM methodology

QM/MM calculations were performed using ChemShell,[7] combining Turbomole[8] and DL_POLY.[9] The polarization of the QM region by the enzymatic environment was included *via* an electronic embedding scheme.[10] The QM/MM boundary was treated by the use of hydrogen link atoms with the charge shift model.[11] The QM part was treated with DFT while the MM part was described by the CHARMM force field. Geometry optimizations were carried out using the BP86 functional[12] in combination with the Def2-SVP basis set[13] for all atoms. Some calculations were carried out using the Def2-TZVPP[14] basis set for Fe in order to test the importance of the basis set, but no large difference has been observed in the computed redox potential (only 0.06 V when the QM region is composed of the heme, His-176 and Asp-238). Optimizations include the heme and all atoms within 13 Å from it. All structures were fully optimized without any symmetry restriction. The energies were recomputed by single-point calculations of the QM region surrounded by the optimized MM point charges using the hybrid B3LYP functional,[15] the Stuttgart-Dresden pseudopotential in combination with its adapted basis set for Fe,[16] and the 6-311G** basis set for all other atoms.[17] We used B3LYP for consistency with the large body of QM/MM studies on related P450 enzymes[18]. The basis sets along with the SDD pseudopotential have been validated in our previous studies on small heme models.[22b] Single-point calculations were carried out with the Gaussian 09 program.[19]

The choice of the QM region

Different QM regions were used to calculate the redox potentials of the enzyme. However, the largest QM region will be used as reference to discuss the geometrical parameters. It is shown in Figure 2 of the main paper. This QM region contains all the first and second coordination sphere residues, including the entire heme, the proximal and distal histidines, respectively His-176 and His-47, as well as the aspartate Asp-238 in interaction with His-176, the arginine Arg-43 in interaction with His-47, the phenylalanine residues Phe-46 and Phe-193, and a water molecule in interaction with both the ferryl-oxo unit and His-47. Histidines appear in

the QM part as methylimidazole, Asp-238 as CH_3COO^-, Phenylalanines as toluene and Arg-43 as ethylguanidinium.

Thermodynamic Born-Haber cycle.

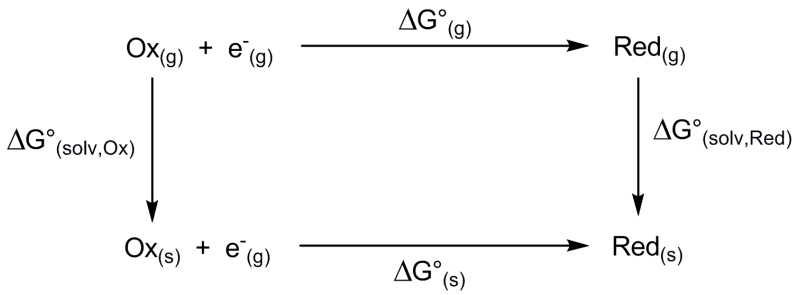

Figure S1. Thermodynamic Born-Haber cycle for the calculation of redox potentials in solution.

$\Delta G°_{(g)}$ is the gas-phase Gibbs free energy of the addition of one electron to the oxidant (i.e. Cpd I in our case) to give the reductant (i.e. Cpd II). It is noteworthy than $\Delta G°_{(g)}$ is the sum of the negative calculated electron affinity, $\Delta E_{(g)}$ (i.e. a negative number if electron uptake is favorable), and the thermal and entropic contributions. This value is usually calculated using quantum mechanics (QM) methods such as density functional theory (DFT). In order to compute the corresponding Gibbs free energy in solution, one has first to calculate the solvation energies of the oxidant and the reductant, respectively $\Delta G°_{(solv,Ox)}$ and $\Delta G°_{(solv,Red)}$ in Figure S1. This is commonly achieved by means of continuum solvation models. $\Delta G°_{(s)}$ is eventually calculated as :

$$\Delta G°_{(s)} = \Delta G°_{(g)} + \Delta G°_{(solv,Red)} - \Delta G°_{(solv,Ox)}$$

The absolute redox potential of the Ox/Red couple $E°_{abs}$ is then defined as :

$$E^0_{abs} = \frac{-\Delta G°(s)}{nF}$$

where n is the number of electrons involved in the half-reaction of interest (here n = 1) and F is the Faraday constant (F = 96485 $J.mol^{-1}.V^{-1}$ or 23.061 $kcal.mol^{-1}.V^{-1}$).

Calculations of redox potentials for model systems.

Absolute redox potentials were calculated for an iron-oxo porphyrin model system with the Born-Haber cycle. The studied molecular system is $[(TMP)Fe(IV)(O)(Im)]^+$/ [(TMP)Fe(IV)(O)(Im)] (with TMP = tetramesitylporphyrin and Im = Imidazole). The redox potential of the Cpd I/Cpd II couple has been determined using three different functionals (B3LYP, BP86 and M06-L), three different basis sets for the small atoms (6-31G**, 6-311G**, 6-311++G**) and three different solvation models (CPCM with UFF radii, CPCM with Pauling radii and SMD). Iron was treated with the Stuttgart-Dresden pseudopotential as well as its adapted basis set. Redox potentials were calculated from both $\Delta G°_{(s)}$ (as shown in the previous section) and $\Delta E_{(s)}$.

Table S1. Calculated absolute redox potentials of $[(TMP)Fe(IV)(O)(Im)]^+$/ [(TMP)Fe(IV)(O)(Im)] with different methodologies. Values in parentheses are the calculated absolute redox potentials where the thermal and entropic contributions are not taken into account (only electronic energies were used for the calculation).

	CPCM (UFF)	CPCM (Pauling)	SMD
B3LYP / B1	4.94 (4.98)	4.67 (4.71)	4.69 (4.73)
B3LYP / B2	5.17 (5.21)	4.91 (4.94)	4.92 (4.96)
B3LYP / B3	5.23 (5.27)	4.97 (5.00)	4.98 (5.02)
BP86 / B1	5.01 (5.02)	4.81 (4.81)	4.79 (4.79)
BP86 / B2	5.24 (5.24)	5.04 (5.04)	5.02 (5.02)
BP86 / B3	5.29 (5.29)	5.09 (5.09)	5.07 (5.07)
M06-L / B1	4.91 (4.97)	4.67 (4.73)	4.68 (4.74)
M06-L / B2	5.07 (5.12)	4.83 (4.89)	4.84 (4.90)
M06-L / B3	5.08 (5.14)	4.85 (4.90)	4.86 (4.91)

These results show that the difference between redox potentials calculated with $\Delta G°_{(s)}$ and $\Delta E_{(s)}$ is not large. Indeed the largest difference is 0.07 V with M06-L. There is no difference at all with BP86.

In the context of the redox potential of the enzyme, the solvation steps of the Born-Haber cycle become redundant because both the protein environment and the explicit solvent molecules of the MM domain polarize the active site in the QM region. The calculation of Gibbs free energies becomes unfortunately quite involved for such systems.[20] Fortunately, test

calculations on heme models show that the difference between redox potentials calculated with electronic energies $\Delta E_{(s)}$ and Gibbs free energies $\Delta G°_{(s)}$ is only around 0.05 V or less (see Table S1). The calculated redox potentials will be given relative to the calculated absolute redox potential of ferrocene, as recommended by IUPAC[21], a procedure that has also been put forward to facilitate error compensation[22].

The choice of spin states

In Cpd I, there is a positively charged porphyrin ring in addition to the iron with its formal oxidation state Fe (IV). As shown in Figure S2, this configuration can lead to a quartet (S = 3/2) or a doublet (S = ½) spin state depending on the spin of the single electron located in the porphyrin orbital (a_{1u} or a_{2u}, depending on the nature of the enzyme). For the well-known P450 enzyme, both spin states are very close in energy[23]. In Cpd II, the heme has been reduced once and the additional electron is located in the orbital of the porphyrin ring. This configuration leads to a triplet (S = 1) spin state. QM/MM calculations were carried out with a simple QM region only composed of the heme unit and His-176. Cpd I was optimized with quartet and doublet spin states. Both states are found to be effectively isoenergetic, with a very slight preference for the quartet (0.8 kcal/mol). Thus all optimizations were carried out with a quartet spin state for Cpd I. Spin contamination was small to negligible in all cases, with $\langle \hat{S}^2 \rangle$ values typically around 2.03-2.04 and 3.79-3.80 for triplet and quartet states, respectively (B3LYP level, slightly lower at BP86).

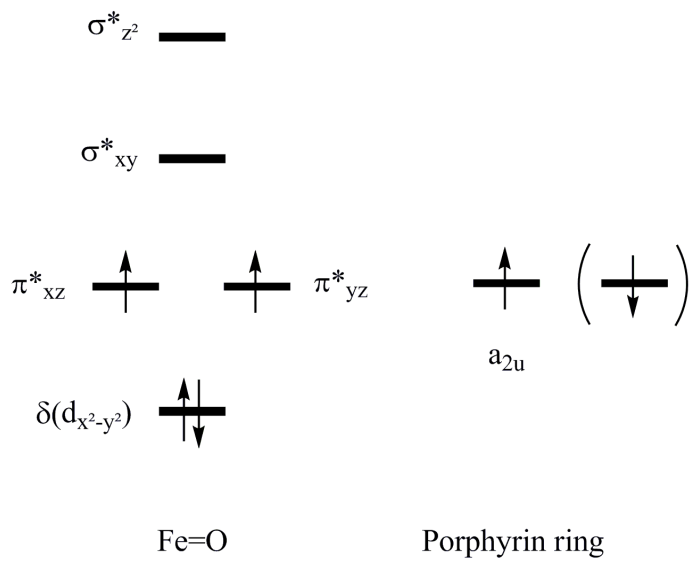

Figure S2. Orbitals and their occupation in Cpd I.

The calculation of the absolute redox potential of ferrocene

As ferrocene is used as the reference redox couple of the study, the absolute redox potential of ferrocene has to be computed at the same level of theory as the redox potential of the enzyme, as shown previously by Konezny *et al.* and Castro *et al.*[22] Thus, ferrocene and ferricinium were both optimized in gas phase at the BP86/Def2-SVP level using Turbomole. Single-point calculations were then carried out with Gaussian 09 and the B3LYP/SDD(Fe)/6-311G** level of theory. Solvation energies were evaluated by a self-consistent reaction field (SCRF)[24] approach based on accurate numerical solutions of the Poisson-Boltzmann equation.[25] Solvation calculations were carried out at the optimized gas phase geometries with the dielectric constant of water and the SMD model was used with its default settings.[26] The absolute redox potential was calculated with the use of the Born-Haber cycle presented in the figure S1 of the SI, affording a value of 5.03 V.[27]

To facilitate repeating the calculations, a sample input for the ChemShell optimization is reproduced below, and a pdb file of compound I is provided as well (largest QM region that started from the 1000 ps equilibration step).

```
#Begin of ChemShell sample input
############################
# O P T I M I Z A T I O N #
############################
# GGA #

global sys_name_id
set dir .

set sys_name_id enzy_prod
global chemsh_default_connectivity_toler
global chemsh_default_connectivity_scale
set chemsh_default_connectivity_toler  0.6
set chemsh_default_connectivity_scale  1.0

# CALL TOPPAR
set top { ./top_all27_prot_na.rtf }
set prm  ./par_all27_prot_na.prm

# SOURCE .CHM with CONNECTIVITY
source /home/lcastro/charmm_1LGA/Largest/save_${sys_name_id}.chm

# load the connectivity from the psf file
load_connect_from_psf  /home/lcastro/charmm_1LGA/Largest/enzy_prod.c
/home/lcastro/charmm_1LGA/Largest/${sys_name_id}.psf

# define active region
source  /home/lcastro/charmm_1LGA/Largest/pdbresidues
source  /home/lcastro/charmm_1LGA/Largest/act
source  /home/lcastro/charmm_1LGA/Largest/qmatoms
```

```
# define residues
set residues $pdbresidues


#QM/MM Optimization
dl-find coords=${dir}/input.c \
        theory=hybrid : [ list \
                        coupling= shift \
                        groups = $groups \
                        cutoff=1000.0 \
                        atom_charges= $charges \
                        qm_region= $qmatoms \
                        debug=no \
                        qm_theory= turbomole : [ list  \
                                                 hamiltonian= bp86 \
                                                 read_control= yes \
                                                 use_ri= yes \
                                                 restart=yes \
                                                 maxcyc=2500 \
                                                 charge=0 \
                                                 mult=4 ] \
                        mm_theory=dl_poly : [ list \
                                              list_option=none \
                                              debug=no \
                                              exact_srf=yes \
                                              use_pairlist=no \
                                              mxlist=16000 \
                                              cutoff=1000 \
                                              mxexcl=2000  \
                                              debug_memory=no \
                                              scale14 = { 1.0 1.0 } \
                                              atom_types= $types \
                                              use_charmm_psf=yes \
                                              charmm_psf_file=
/home/lcastro/charmm_1LGA/Largest/${sys_name_id}.psf \
                                              charmm_parameter_file= $prm \
                                              charmm_mass_file= $top ] ] \
        coordinates=hdlc \
        result= ${dir}/bp86-reactant1.c \
        residues= $residues \
        active_atoms= $act \
        optimiser=lbfgs lbfgs_mem=50 trustradius=const \
        maxstep=0.9 list_option=medium \
        maxcycle=700 maxene=3000 tolerance=0.00045


times
#End of ChemShell sample input
```

Additional Results

1. Geometries.

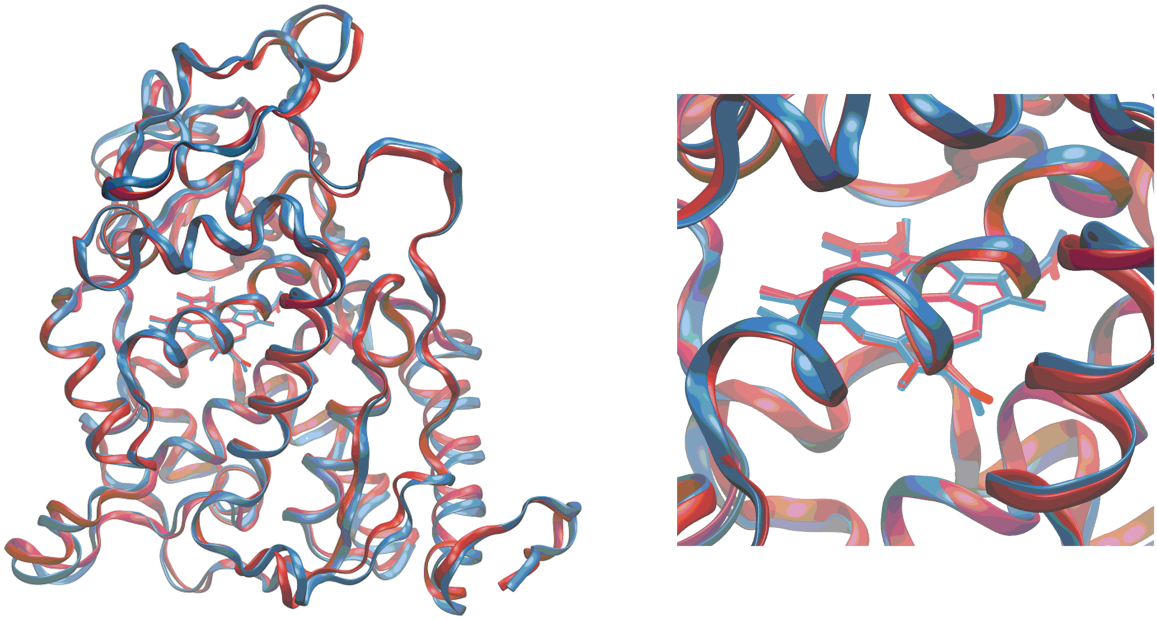

Figure S3. Left: Superposition of the optimized structure of Cpd I (in blue) and the experimental structure of the resting state (in red). A zoom into the active site is depicted on the right.

Calcium Binding Sites. Two calcium cations maintain the global integrity of LiP. One is located in the proximal region of the enzyme whereas the other one is located in the distal region. The global configuration of the residues around each calcium seen in the experimental structure of the resting state is preserved during optimization of Cpd I.

Protein-Heme Interactions. The heme interacts with the protein mainly *via* three different ways. The first important interaction is between the iron atom and the ε nitrogen of the proximal histidine, His-176. The Fe-N distance is found to be 2.14 Å, which is slightly larger than the unrestrained experimental distance for the resting state (2.05 Å).[1] This suggests that the Fe-N bond is weakened by the *trans* effect of the additional oxo group in Cpd I. Secondly, the "outer" propionate (see Figure 2) substituent of the porphyrin ring is protonated in order to form an H-bond with one carboxyl oxygen atom of Asp-183. The H---O(=C) bond is calculated to be 1.46 Å. The "inner" propionate only forms H-bonds with water molecules as an acceptor. Poulos *et al.* suggested that one H-bond may exist between the "inner" propionate and the peptide NH of Ala-180,[1] but the optimization leads to a O---H distance of

2.26 Å, which is too long for a strong H-bond. The third interaction is indirect and surprising. The ferryl oxo group forms an H-bond with one hydrogen of a water molecule, which forms another donor-interaction with the ε nitrogen of the distal histidine, His-47. The water molecule thus donates two H-bonds (and accepts one from another water molecule). This bridging water molecule can be seen in Figure 2. This observation is in disagreement with the proposal of Poulos *et al.* that the ferryl oxygen of Cpd I is stabilized by an H-bond with the Arg-43 residue.[1] Instead, Arg-43 forms a donor H-bond with the ε nitrogen of the distal histidine, His-47.

Differences between Cpd I and Cpd II. The minor differences that exist between the optimized structures of Cpd I and Cpd II mostly appear in the active site of the enzyme. Relevant bond lengths are presented in Table 1. The addition of one electron into the a_2u orbital of the porphyrin ring does not affect the iron-oxo bond, but pulls the hydrogen atom from the bridging water molecule toward the oxo group (from 1.71 Å to 1.68 Å) due to reduced repulsion between the positively charged hydrogen and the now neutral porphyrin moiety. As a consequence, the H-bond length between water and His-47 slightly increases (from 1.86 Å to 1.91 Å). Upon reduction, the Fe-N bond becomes longer, as well as the distance between His-176 and Asp-238, because of electrostatic repulsion with the additional electron on the porphyrin ring. All these differences are quite small though, never exceeding 0.05 Å.

Table S2. Relevant calculated bond lengths of the active site for Cpd I and Cpd II.

	Cpd I	Cpd II
Fe=O	1.66	1.66
FeO-H_2O	1.71	1.68
H_2O-N (His-47)	1.86	1.91
Fe-N (His-176)	2.14	2.17
H (His-176) – O (Asp-238)	1.77	1.82

The optimized geometrical parameters show clearly that no dramatic structural changes appear during the catalytic cycle of lignin peroxidase. The resting state, Cpd I and Cpd II all have the same global features and the only significant change induced by the additional oxo group is the polarization of a water molecule.

2. Redox potentials.

2.1 Neutral wild type

In order to see if BP86 and B3LYP have the same behavior, absolute redox potentials (ARP) calculated with BP86 have been plotted versus the ones calculated with B3LYP. The plot is shown in Figure S4. The absolute values differ but the trend is the same, with a slope almost equal to 1. It is therefore justifiable to use either BP86 (as recommended for some compound I models),[22b] or B3LYP (as widely used in QM/MM calculations of P450 species).

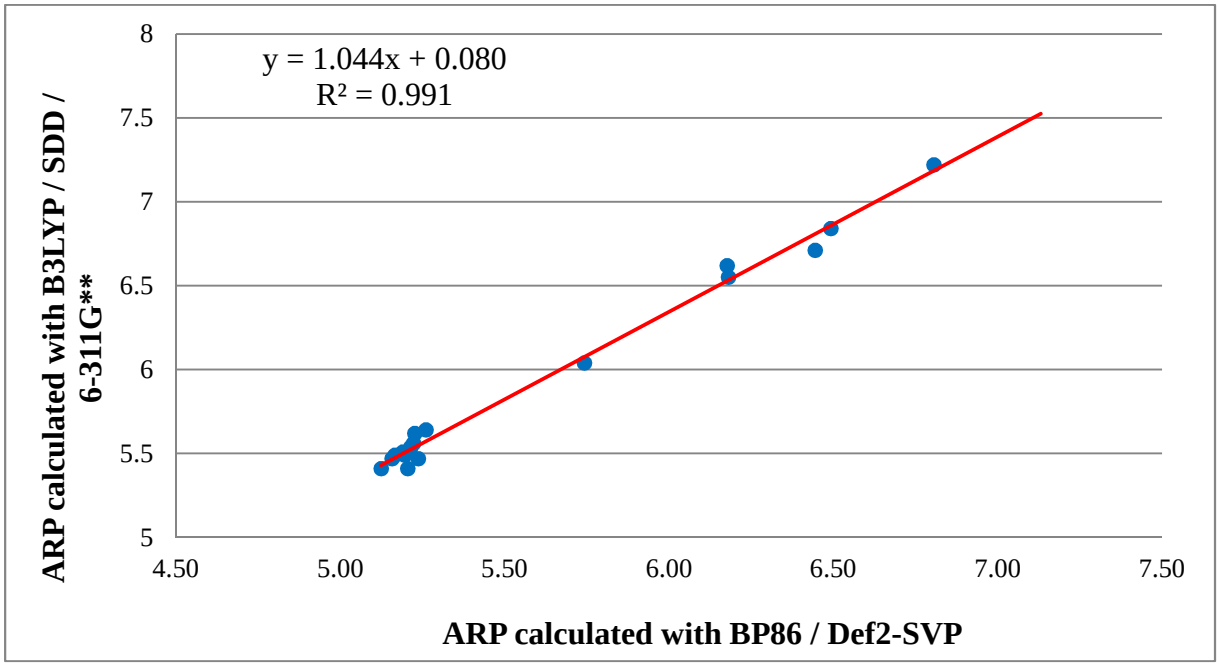

Figure S4. Plot of the ARP calculated with BP86/Def2-SVP versus the ARP calculated with B3LYP/SDD/6-311G**. The points correspond to a variety of situations (different QM regions or systems with protonated residues (see text)).

The relative redox potentials (RRPs) calculated with different QM regions are given in Table S3. The smallest QM region is composed of the heme (the iron-oxo moiety and the porphyrin ring with all its substituents) and the proximal histidine His-176, directly bonded to the iron center. The RRP is calculated to be 0.38 V. If Asp-238 is included inside the QM region, the RRP remains exactly the same. This means that both the H-bond between His-176 and Asp-238 and the negative charge of aspartate are well described by the MM region in the QM/MM calculation. The electrostatic interaction between this negative charge and the heme can be as well modeled by the QM/MM interaction as by a full QM description. The addition of Arg-43 shows how the electrostatic interaction between the positive charge and the heme is handled

by the QM/MM interaction. The RRP is equal to 0.44 V, which differs only by 0.06 V from the smallest QM region. The additional Phe residues inside the QM region do not really change the RRP, meaning either that the π-π interactions between the Phe residues and other aromatic rings (His-176 in particular) are fairly well described by the QM/MM interaction, or that they do not play any role in the value of the RRP. The RRP calculated when the distal histidine His-47 and the water molecule are present in the QM region is equal to 0.46 V. This shows that these bridging water H-bonds are not perfectly accounted for at the MM level. The difference with respect to the smallest region is only 0.08 V, though. The RRP calculated for the largest QM region shown in Figure 2 is 0.48 V.

Table S3. Relative redox potentials (V versus $FeCp_2^+/FeCp_2$ in water) calculated with different sizes of QM regions.

QM region	**RRP (V / $FeCp_2^+/FeCp_2$)**
A. Heme + His-176	0.38
B. Heme + His-176 + Asp-238	0.38
C. Heme + His-176 + Arg-43	0.44
D. Heme + His-176 + Asp-238 + Arg-43	0.47
E. Heme + His-176 + Asp-238 + Arg-43 + H_2O	0.51
F. Heme + His-176 + Asp-238 + Phe-46 + H_2O	0.53
G. Heme + His-176 + Asp-238 + Phe-193 + H_2O	0.46
H. Heme + His-176 + Asp-238 + His-47 + H_2O	0.46
Full. Heme + His-176 + Asp-238 + Arg-43 + Phe-46 + Phe-193 + His-47 + H_2O	0.48

2.2 Protonated wild type

In order to gain some insights about the relationship between the activity of the enzyme and the pH, we performed additional QM/MM calculations for systems where specific residues have been protonated. In order to maintain the global neutral charge of the enzyme, one proton has been systematically removed from a protonated aspartate or a protonated glutamate located on the surface of the enzyme, far from the active site.

In the first system, Asp-238 has been protonated (Aspp-238) and the QM region is composed of the heme, His-176 and Aspp-238 (equivalent to QM region **B**). The optimized structure of

Cpd I is shown in Figure S5a. His-176 still forms a H-bond with the carboxyl oxygen of Aspp-238, but the O-H distance is 2.40 Å, i.e. much longer than the H-bond when Asp-238 is not protonated (1.80 Å for the QM region **B**). This suggests that this H-bond is dramatically weakened after protonation of the residue and that the hydrogen atom is completely located on the histidine rather than shared between His-176 and Aspp-238. Goodin *et al.* proposed that the position of the hydrogen atom between the nitrogen of the histidine and the oxygen of aspartate dictates the degree of imidazolate character of the proximal histidine for Cytochrome *c* Peroxidase.[28] A strong imidazolate character of the histidine residue should have an impact on the redox potential of the enzyme since the "negative" ligand should stabilize the positive charge of Cpd I, and thus lower the redox potential. Here, when Asp-238 is not protonated, the N-H bond length is 1.07 Å. When Asp-238 is protonated, it is 1.03 Å; i.e. the decrease upon protonation is very small. The Fe-N distance is equal to 2.16 Å when Asp-238 is protonated, which is nearly the same as when it is not (2.15 Å). These comparisons suggest that the strength of the H-bond between His-176 and Asp-238 has no real impact on the interaction of His-176 with the iron center because the hydrogen atom is never shared between both residues. In order to verify this, we attempted to optimize the original unprotonated system where the H has been moved manually from His-176 to Asp-238, i.e. the latter becoming protonated and the former a negatively charged imidazole. During the optimization, the hydrogen atom directly goes back to His-176, suggesting that this hydrogen indeed is never shared between both residues. Upon protonation, the RRP is predicted to increase from 0.38 V to 1.52 V. This is a very strong increase which can be explained by the fact that the neutralization of the negative charge of Asp-238 removes a stabilizing electrostatic attraction between this residue and the positive charge of the porphyrin ring. The protonation leads to the destabilization of Cpd I and a large increase of the redox potential.

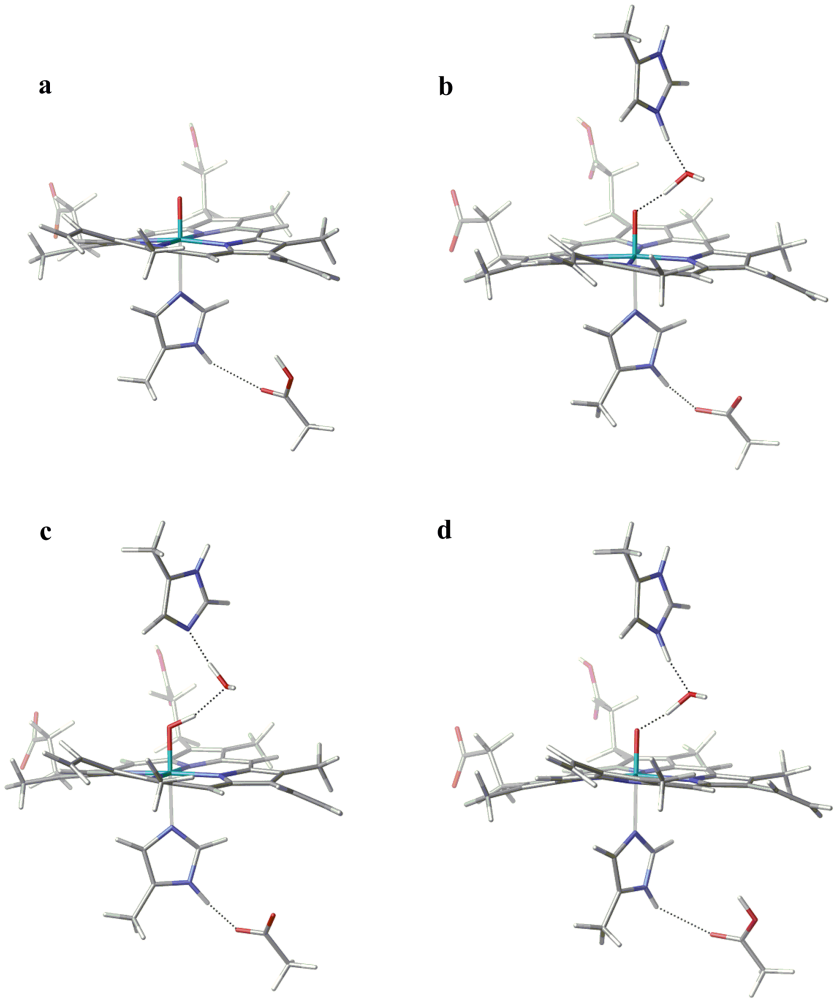

Figure S5. Optimized QM/MM structures of the QM region only for : a) Cpd I when Asp-238 is protonated, b) Cpd I when His-47 is protonated, c) Cpd II when His-47 is protonated, d) Cpd I when both Asp-238 and His-47 are protonated.

Another candidate for protonation at low pH is His-47. To probe this, the heme, His-176, Asp-238, the protonated His-47 (Hsp-47) and the water molecule have been included into the QM region (as in QM region **H**). The optimized structures of Cpd I and Cpd II are presented in Figure S5b and S5c respectively. In Cpd I, the orientation of the water molecule changes because it becomes an acceptor for an H-bond with His-47. The liberated hydrogen atom of water is dangling, i.e. not involved in other interactions. As a consequence, the remaining H-bond with the oxo group is strengthened (1.35 Å versus 1.75 Å in **H**) and the Fe=O bond is weakened (1.68 Å versus 1.65 Å in **H**). The *trans* effect towards His-176 becomes smaller and thus the Fe-N shortens (2.12 Å versus 2.15 Å in **H**). In Cpd II, two concerted hydrogen transfers take place during the optimization. The additional hydrogen atom located on the nitrogen of His-47 is transferred to the water molecule while one hydrogen atom of the water

molecule is transferred to the ferryl oxo group. As a consequence, Cpd II does not present a ferryl oxo group anymore, but a ferryl hydroxo. The Fe-O distance becomes 1.76 Å. Because of the much weakened *trans* effect, the Fe-N distance shortens to 2.08 Å. The reason for this concerted transfer is the filling of the a_2u orbital of the porphyrin ring in Cpd II. In Cpd I, the positive charge on the ring prevents the transfer of the H of water because of a slight electrostatic repulsion. Upon protonation of His-47, the RRP is calculated to increase from 0.46 V to 1.68 V. For the same reason as for the previous system, Cpd I is destabilized by the close proximity between the protonated residue and the electron-deficient porphyrin, and the redox potential increases. When both Asp-238 and His-47 are protonated (see the Cpd I structure in Figure 6d), the RRP further increases to 2.19 V. Thus, electrostatic repulsion seems to be the key for the destabilization of Cpd I with respect to Cpd II and the concomitant increase of the redox potential under low pH conditions.

2.3 Dynamic averaging

In order to check the importance of the starting snapshot for the QM/MM calculations, we calculated the redox potentials for two different protonation states (the "neutral wild type" Cpd I/Cpd II system and the situation where Asp-238 and His-47 are both protonated) by performing the QM/MM optimizations for additional snapshots, taken every 100 ps between 100-900 ps for the wild type, and every 200 ps between 200-800 ps for the protonated version. Redox potentials were calculated with the same QM/MM methodology and single-point B3LYP/6-311G**/SDD calculations were carried out as before. An overlap of the Cpd I geometries optimized from the five different snapshots is presented in Figure S6. The backbone of the enzyme is basically the same during the entire molecular dynamics simulation, but the side chains move significantly. These movements induce a large variation of the distance between the iron atom and the charged residues during the simulation that leads to quite different calculated redox potentials (see Table 3). The standard deviations are indeed quite high (0.33 V and 0.45 V for the "neutral wild type" and the protonated variant, respectively). The values calculated for the Aspp-238 / Hsp-47 situation are especially large, since they can exceed 3 V with respect to the ferrocene couple. This situation is extreme because the nearest charged residues (and thus the most internal ones) have been protonated. A more realistic situation at pH 3 should lie between these two states.

Table S4. Redox potentials calculated for two different protonation states, given in V with respect to the ferrocene couple. QM/MM calculations were carried out on the starting geometries of five or ten different snapshots.

Snapshot	"neutral wild type"	Aspp-238 / Hsp-47
100 ps	0.79	
200 ps	0.79	2.62
300 ps	0.80	
400 ps	1.08	3.11
500 ps	0.01	
600 ps	0.33	2.27
700 ps	1.16	
800 ps	0.72	3.31
900 ps	0.66	
1000 ps	0.38	2.19
Average (standard dev.)	*0.67(33)*	*2.70(45)*

The average calculated for the "neutral wild type" is 0.67 V with respect to the ferrocene redox couple, which means that the absolute redox potential (ARP) is equal to 5.70 V. The estimated experimental value of the redox potential of LiP is 1.2 V/SHE. [29] Since the experimental value for the standard hydrogen electrode in water is 4.36 V, the estimated experimental ARP of LiP is 5.56 V, which is very close to the theoretical value. This good agreement may to some extent be fortuitous, for a number of reasons: we are comparing the value for the neutral wild type with an experimental estimate at low pH, the latter of which is associated with a large uncertainty, and for the former, more snapshots would arguably be needed in order to have a more precise average. These potential sources of error notwithstanding, the comparison is promising and shows that the computed value is certainly in the right order of magnitude. In practice we are more interested in relative trends in redox potentials, and QM/MM studies may be a useful tool to predict such trends across proteins and enzymes and to obtain insights about their reductive or oxidative properties. In the following we explore such trends to design mutants with bespoke redox potentials.

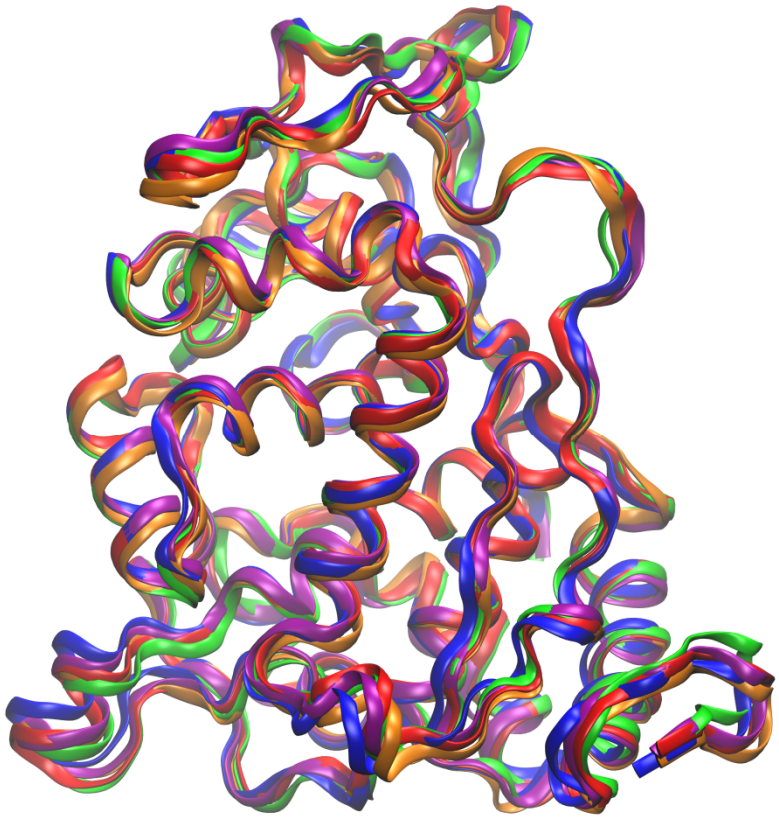

Figure S6. Overlap of the Cpd I backbone geometries optimized from five different snapshots (200 ps in blue, 400 ps in green, 600 ps in red, 800 ps in orange, 1000 ps in purple).

Redox potentials after deletion of the charges of some residues.

Table S5. Calculated B3LYP RRP for an optimized system after deletion of the charges of specific neutral residues.

Deleted charges	Distance from Fe (Å)	RRP (V / $FeCp_2^+/FeCp_2$)
His-47	5.2	0.43
Water	2.9	0.27
Phe-193	5.5	0.23
Phe-46	5.3	0.19
Ile-235	6.5	0.24
Met-172	9.2	0.39

Table S6. Calculated B3LYP RRP for an optimized system after deletion of the charges of specific negatively charged residues.

Deleted charges	Distance from Fe (Å)	RRP (V / $FeCp_2^+$/$FeCp_2$)
Asp-238	6.64	2.24
Asp-183	10.9	1.64
Asp-194	12.6	1.39
Asp-141	14.4	1.22
Asp-48	14.4	1.16
Asp-201	14.9	1.21
Asp-68	17.6	1.03
Asp-264	18.0	1.06
Asp-244	18.7	1.04
Asp-160	18.9	1.01
Asp-165	18.9	1.10
Asp-75	20.7	0.96
Asp-107	22.3	0.92
Asp-284	27.0	0.80
Glu-40	12.0	1.59
Glu-146	12.2	1.39
Glu-78	14.6	1.19
Glu-224	15.4	1.20
Glu-207	15.7	1.12
Glu-37	16.7	1.13
Glu-250	19.0	1.02

Table S7. Calculated B3LYP RRP for an optimized system after deletion of the charges of specific positively charged residues.

Deleted charges	Distance from Fe (Å)	RRP (V / $FeCp_2^+$/$FeCp_2$)
Arg-43	5.8	-1.73
Arg-157	15.2	-0.68
Arg-234	16.0	-0.62
Arg-243	17.6	-0.54
Arg-246	19.5	-0.46
Arg-132	19.6	-0.44
Arg-337	20.3	-0.48
Arg-211	25.2	-0.28
Lys-260	18.9	-0.44
Lys-293	21.0	-0.36
Lys-95	22.9	-0.33
Lys-308	23.1	-0.33
Lys-342	26.0	-0.28
Lys-133	28.7	-0.18
Lys-59	29.7	-0.20
Lys-100	30.2	-0.19
Lys-61	31.3	-0.16
Lys-7	32.5	-0.15

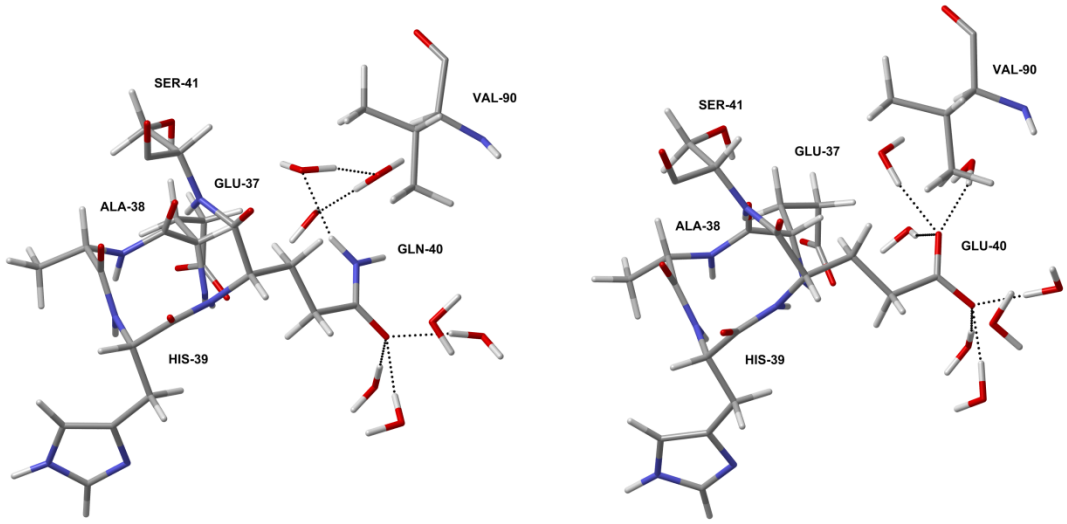

Figure S7. Optimized local structures of Cpd I (left) and the E40Q mutant (right) around residue 40.

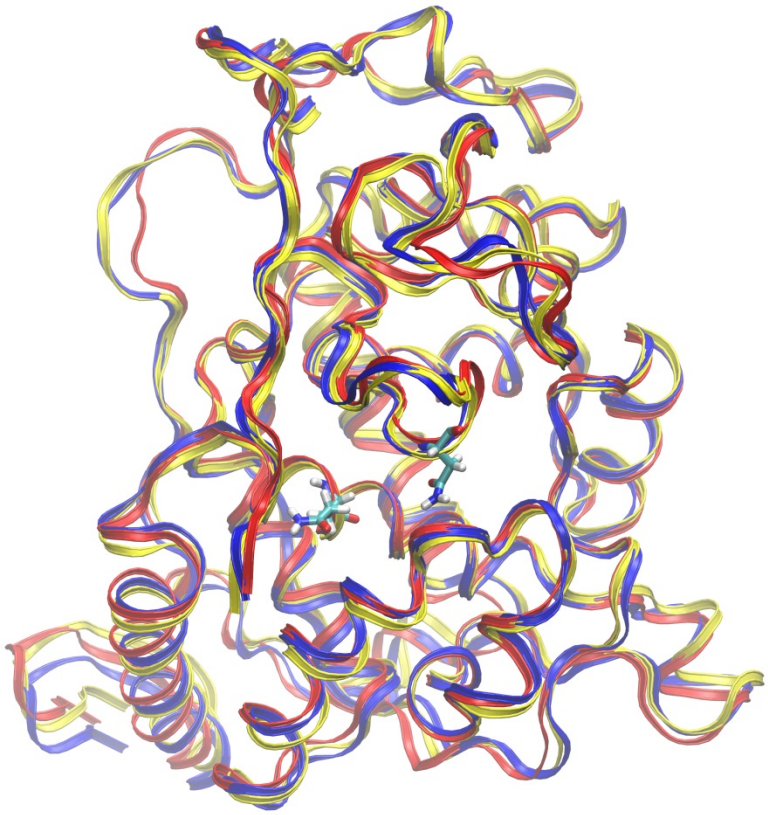

Figure S8: Overlay of the MM-equilibrated structures of compound I of the wild type (in yellow) and the E40Q and D183N mutants in red and blue, respectively (mutated residues shown explicitly).

References

1 T. L. Poulos, S. L. Edwards, H. Wariishi, M. H. Gold, *J. Biol. Chem.*, 1993, **268**, 4469-4440.

2 Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins : Struct., Funct., Bioinf.*, **2005**, *61*, 704-721.

3 This charge neutralization was done in order to avoid artifacts from spurious long-range electrostatic interactions. In reality the system must be charge-neutral, so any excess charges (including that of an added proton close to the active site) will be balanced by counterions in the bulk solution. The most likely location for counter-anions, for example, is close to a positively charged residue at the periphery. Rather than modelling such contact- or solvent-separated ion pairs at the periphery, we just neutralize a protonated residue by removing the proton.

4 Brunger, A. T.; Karplus, M. *Proteins : Struct., Funct., Genet.*, **1988**, *4*, 148-156.

5 Brooks, B. R.; Brooks III, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comp. Chem.*, **2009**, *30*, 1545-1615.

6 (a) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.*, **1983**, *4*, 187-217. (b) Chu, J.-W.; Trout, B. L.; Brooks, B. R. *J. Chem. Phys.*, **2003**, *119*, 12708-12717.

7 P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjovoll, A. Fahmi, A. Schafer, C. Lennartz, C. *Theochem*, 2003, **632**, 1-28.

8 TURBOMOLE V6.4 2012, a development of University of Karlsruhe and Forschungszentrum Karlsuhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

9 W. Smith, T. R. Forester, *J. Mol. Graphics*, 1996, **14**, 136-141.

10 D. Bakowies, W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580-10594.

11 A. H. de Vries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto, G. J. Kramer, *J. Phys. Chem. B*, 1999, **103**, 6133-6141.

12 (a) A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098-3100. (b) J. P. Perdew, *Phys. Rev. B*, 1986, **33**, 8822-8824.

13 (a) F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305. (b) A. Schaefer, H. Horn, R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571-2577.

14 A. Schaefer, C. Huber, R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829-5835.

15 (a) A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648-5652. (b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.

16 M. Dolg, U. Wedig, H. Stoll, H. Preuss, *J. Chem. Phys.*, 1987, **86**, 866-872.

17 W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257-2261.

18 (a) A. Altun, D. Kumar, F. Neese, W. Thiel, *J. Phys. Chem. A*, 2008, **112**, 12904-12910. (b) K.-B. Cho, H. Hirao, H. Chen, M. A. Carvajal, S. Cohen, E. Derat, W. Thiel, S. Shaik, *J. Phys. Chem. A.*, 2008, **112**, 13128-13138. (c) D. Fishelovitch, C. Hazan, H. Hirao, H. J. Wolfson, R. Nussinov, S. Shaik, *J. Phys. Chem. B*, 2007, **111**, 13822-13832. (d) J. C. Schöneboom, H. Lin, N. Reuter, W. Thiel, S. Cohen, F. Ogliaro, S. Shaik, *J. Am. Chem. Soc.*,

2002, **124**, 8142-8151. (e) J. C. Schöneboom, S. Cohen, H. Lin, S. Shaik, W. Thiel, *J. Am. Chem. Soc.*, 2004, **126**, 4017-4034.

19 Gaussian 09, Revision A.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmaverni, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

[20] Such methods, e.g. based on free-energy perturbation techniques, have been used to calculate redox potentials of enzymes, see e.g.: J. Heimdal, M. Kaukonen, M. Srnec, L. Rulisek, U. Ryde, *ChemPhysChem*, 2011, **12**, 3337-3347.

21 G. Gritzner, J. Kuta, *Pure Appl. Chem.*, 1984, **56**, 461−466.

22 (a) S. J. Konezny, M. D. Doherty, O. R. Luca, R. H. Crabtree, G. L. Soloveichik, V. S. Batista, *J. Phys. Chem. C*, 2012, **116**, 6349-6356. (b) L. Castro, M. Bühl, *J. Chem. Theo. Comput.*, 2014, **10**, 243-251.

23 S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, *Chem. Rev.*, 2005, **105**, 2279-2328.

24 (a) S. R. Edinger, C. Cortis, P. S. Shenkin, R. A. Friesner, *J. Phys. Chem. B*, 1997, **101**, 1190-1197. (b) M. Friedrichs, R. H. Zhou, S. R. Edinger, R. A. Friesner, *J. Phys. Chem. B*, 1999, **103**, 3057-3061. (c) B. Marten, K. Kim, C. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnalda, D. Sitkoff, B. Honig, *J. Phys. Chem.*, 1996, **100**, 11775-11788.

25 A. A. Rashin, B. Honig, *J. Phys. Chem.*, 1985, **89**, 5588-5593.

26 A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378-6396.

27 This value compares reasonably well to previous data in the literature, cf. 4.92 V at B3LYP/SDD/6-311G**(SMD,CH_2Cl_2, reference 22b), or 5.24 V at B3LYP/SDD/6-311G**(PCM,DMSO), calculated from the data in: L. E. Roy, E. R. Batista, P. Hay, *J. Inorg. Chem.* **2008**, *47*, 9228-9237.

28 D. B. Goodin, D. E. McRee, *Biochemistry*, 1993, **32**, 3313-3324.

29 P. J. Kersten, B. Kalyarnaraman, K. E. Hammel, B. Reinhammar, *Biochem. J.*, 1990, **268**, 475-480.